00001006S

CONTAMINATION OF SURFACE SOIL IN COLORADO BY PLUTONIUM, 1970 - 1989:

SUMMARY AND COMPARISON OF PLUTONIUM

CONCENTRATIONS IN SOIL IN THE ROCKY FLATS PLANT VICINITY AND EASTERN COLORADO

By Robert W. Terry, Colorado Department of Health

The Colorado Department of Health has studied plutonium contamination in surface soil in 13 areas to a distance of 11.3 km (7 miles) from the center of the Rocky Flats Plant and in 8 areas in Eastern Colorado. Eleven surveys have been conducted from 1970 to 1989. An identifiable plutonium concentration gradient can be observed in the Rocky Flats plant vicinity, with the highest concentrations closest to the Rocky Flats Plant. Higher concentrations of plutonium may be found ESE of the plant than in any other direction, consistent with the local pattern of wind, which typically blows across Rocky Flats from the west and northwest. Within each sampling area, measured plutonium concentrations have declined steadily since the first survey was conducted in 1970. Measured plutonium concentrations are grouped by sampling area and are therefore grouped with respect to distance and direction from the center of the Rocky Flats Plant. Regression analysis of plutonium concentrations against time for each sampling area is presented, with confidence intervals on the regressions. Assuming that the current trend continues, plutonium concentrations in the top 0.64 cm (0.25 inch) of soil can be expected to fall below 0.037 Bq/gm (1 pCi/gm) in all areas outside the Rocky Flats Plant boundary by 1995.

INTRODUCTION

Several studies have been made of plutonium contamination in the Rocky Flats Plant vicinity over the years. Most of these studies have been either one-time evaluations or reviews of information that was gathered from various sources. In nearly all cases the observed plutonium contamination around Rocky Flats and elsewhere appears to have been viewed as a static situation, rather than a dynamic one where plutonium concentrations can change not only with respect to the distance and direction from the Rocky Flats Plant, but over time as well.

*This project was supported primarily by the General Fund of the State of Colorado. A substantial contribution to this project was funded by a Mutual Cooperation Agreement, together with Amendments No. 1, No. 2, No. 3 and No. 4, between the United States Department of Energy and the State of Colorado.

Central Rocky Mountain Chapter/Health Physics Society
Annual Technical Meeting, April 22, 1991
CONTAMINATION OF SURFACE SOIL IN COLORADO BY PLUTONIUM, 1970 - 1989

This report provides a review of a unique data set. The Colorado Department of Health Surface Soil Survey has provided sensitive and precise measurements of $^{239+240}$Pu in 11 separate years over a 20-year period. The methods of sample collection and analysis have remained essentially unchanged during that entire period.

The purposes of this report are

1. to support the existing body of conclusions that clear and identifiable concentration gradients of plutonium exist in the vicinity of the Rocky Flats Plant, of increasing $^{239+240}$Pu concentration as one approaches the Plant site, are present;
2. to demonstrate that concentrations are changing, more or less systematically over time, and are generally falling; and
3. to fill in gaps in the body of data, for years in which the survey was not conducted, or, in years when the survey was conducted, to estimate the concentrations in areas where measurements were not completed.

The data set will eventually be presented in a pictorial format, with concentration isopleth lines drawn on a separate map for each year. For that format, the calculated locations of each point on the isoconcentration lines will be strengthened by complete data sets. Maps showing projected patterns of distribution of the contamination in future years will also be drawn.

METHODS

The overall design of the survey, method of sample collection and measurement techniques have been described elsewhere (Co1977). This report relies on the measurement results in their present form for statistical analysis of the data sets. While the entire data set is characterized by exceptional continuity, minor modifications in procedures may have had an effect on the results. The reader is encouraged to bear this fact in mind throughout this report.

Each data set is the historical record of $^{239+240}Pu$ measurements within a specific area, or sector. Grouping the data by sector serves to isolate each data set with respect to distance and direction from the center of the Rocky Flats Plant. The measured concentrations in each sector were then graphed as a dependent function of time. Based on a visual analysis of the data, a decision was made as to the type of regression analysis that would be performed. In almost all cases the selected method of regression analysis was an exponential fit. Exponential fit was generally the most desirable choice for the following reasons:

1. Regardless of whether the concentrations appeared to increase over time or decrease, the fit would not in any case estimate plutonium concentrations to be less than zero; a linear regression would do so.
2. The measured concentrations for the most part appeared to be declining over time, approaching zero in a manner consistent with an exponential model.
3. High-order polynomial analysis can be used to force a regression line through every data point, without regard to proposing a mechanism that would explain all of the various coefficients that are generated; it is an extreme form of empirical science.
4. Exponential fit provides a way of understanding a mechanism that causes the measured plutonium concentrations to decline in a way that is largely independent of any parameter besides time.

In three cases a linear regression was used for part of the data set because an exponential fit did not seem to adequately describe the subjective observation of the pattern of variation in the data. In all three cases the data set was divided into either two or three subsets. Because the data sets are all very small, consisting of no more than 11 elements, regression analysis of the subsets left very few degrees of freedom.

The effect of small data sets on the reliability of conclusions that are drawn from regression analysis cannot be overlooked. Therefore, all of the regression analysis presented here relies on confidence intervals to clearly describe the uncertainty of the estimated regressions.

While simple (unweighted) linear regression is straightforward, it should be constructive to describe the formulas that were used in this analysis. First, the equation describing the relationship between $^{239+240}Pu$ concentration (Y) and time (X) takes the following form:

$$Y = a + bX$$

or in the case of an exponential fit, the regression postulates a direct linear relationship not between X and Y but between X and the *natural logarithm of* Y:

$$\ln Y = a + bX$$

$$\text{or} \quad Y = e^{a + bX} = e^{a} \cdot e^{bX} = ke^{bX}.$$

The final form for the exponential fit, $Y = ke^{bX}$, is probably the most widely used. However, in order to provide a simple understanding of the regression technique this report will rely on the first form,

$$Y = e^{a + bX},$$

to provide a direct calculation of Y, the estimated $^{239+240}Pu$ concentration in a sector, as a function of X, the year.

Whether using Y or lnY, both are treated as Y in the paradigm for calculating the regression coefficients, as follows (P11981):

n = number of X,Y pairs employed in the regression

SST (sum of squares, total [Y]) = $\Sigma(Y - \bar{Y})^2$

SCP (sum of cross products of X and Y) = $\Sigma(X - \bar{X})(Y - \bar{Y})$

SSX (sum of squares, total [X]) = $\Sigma(X - \bar{X})^2$

$b = SCP/SSX$

$a = \bar{Y} - b\bar{X}$

$\bar{Y}_c = a + bX$

$$SSR = b \cdot SCP$$

$$SSE = SST - SSR$$

$$r^2 = SSR/SST$$

and $$s_{YX} = [SSE/(n - 2)]^{1/2}.$$

Finally, confidence bands about the regression line are defined as the set of confidence intervals on each estimated value of Y, or $\overline{Y}_c$:

$$\overline{Y}_c \pm t \cdot s_{YX} . [1 + (1/n) + (X_o - \overline{X})^2/SSX]^{1/2},$$

where t is the t-value for the 95% confidence level.

It should again be emphasized that when lnY is treated as Y in the formulas listed above, ultimately $\overline{Y}_c$ will be transformed to *estimated concentration*, <u>not</u> the estimated *logarithm of concentration*. Similarly, the upper and lower limits of the confidence intervals, or confidence bands on the regression lines, are also transformed as a power of *e*.

In this report, residuals analysis was performed only on the final results of either the linear regression or the exponential fit, by calculating the difference between the estimated concentration for a given year and its corresponding measured concentration (Dr1966):

$$e_i = Y_i - \hat{Y}_i.$$

For the purposes of this report, only a visual inspection of the plotted residuals has been used. Further statistical analysis of the residuals has not been performed.

Only one X,Y data pair was rejected from the body of information that was analyzed. That was a high measured concentration in Sector 10, that was eliminated from the data set because it was known to have resulted from poor conditions at the time of measurement and because it was clearly an outlier. In working with measurements that are near or below the lower limit of detection it is difficult to identify low measured concentrations that should be treated as outliers; the conclusions of this report may be compromised by a failure to identify outliers on the low side.

The Radiation Control Division of the Colorado Department of Health uses a unique convention in the compilation and presentation of measured values that grows out of a strict adherence to the concept of Lower Limit of Detection (LLD). Under no circumstances are best estimates of concentration ever retained in records or reports. Therefore, LLDs have been used in the regression analysis described in this report.

Admittedly, for analysis of this type a measured value that is nothing more than an LLD contains less information than a best estimate of concentration. However, it contains more information than nothing at all. Therefore, the conclusions of this report are somewhat compromised by giving equal weight to both LLDs and actual measured concentrations.

The use of LLDs in the regression analysis is based on two considerations:

1. The data sets are so small that even information of slightly diminished value, such as LLDs, can help to make the results more robust.
2. For the most part, the *actual* $^{239+240}$Pu concentrations are probably not very different from the value of the LLD, and therefore the use of LLDs in the regression probably does not degrade the results severely.

RESULTS AND DISCUSSION

Figures 1 and 2 illustrate the steps used in performing the regression analysis in two of the sectors. Figures 3 and 4 summarize the results of all of the regression analysis that was performed. In three locations, Sectors 2, 5 and 8, there was clearly a discontinuity in the measured $^{239+240}$Pu concentrations. There is little question that using measured concentrations for the years 1970 - 1974 in these three sectors would diminish the accuracy of regression analysis of the measured concentrations from subsequent years. Therefore, these data sets were divided into subsets. For Sector 2, the data point from 1970 was not used in the linear regression analysis that was used to interpolate a value for 1973.

In the areas of highest plutonium concentration, the relatively narrow confidence intervals lead to the conclusion that fairly precise predictions can be made about past and future plutonium contaminations in surface soil around the Rocky Flats Plant. In many cases, the predictive value of the regression is limited, as illustrated by the very wide confidence bands. Although the confidence bands in most of the sectors are relatively wide, it must be emphasized that the absolute magnitude of the values described by the upper confidence bands is generally not exceptionally high. Persons who function in public policy or public planning roles will, it is hoped, be able to confidently base decisions on predictions that plutonium contamination will not in any case exceed specified values, and should be able to evaluate the reliability of these predictions from a review of the compiled graphs as a whole.

Figures 6 and 7 show the results of residuals analysis. For the most part, the residuals are randomly distributed around zero; more simply, the regression lines tend to cut through the middle of the plotted measured values, indicating that no serious biases have been introduced by the regression models and that the regression models are probably not grossly inappropriate.

In general plutonium concentrations appear to be falling in all locations near the Rocky Flats Plant and throughout the State of Colorado. In some cases where the regression analysis estimated a trend toward slightly, but nonetheless increasing concentrations, this deviation from expectations may be due to an imperfect set of data, but does not lead in any case to expectations of alarming concentrations of plutonium in the environment.

Table 1 tabulates the coefficients from all of the regression analyses, together with r^2 (the *coefficient of determination*), s_{YX} (the *standard error of estimate*) and SSE (the sum of squares, error, or total unexplained variation in each data set). The coefficients for the regression analyses, s_{YX} and SSE values are provided to assist in calculation of specific estimated $^{239+240}$Pu concentrations, together with their associated confidence bands. t-values in this report are for the 95% confidence level.

The *coefficient of determination*, or r^2, describes that fraction of the total variation in the measured $^{239+240}$Pu concentrations that is explained by the regression model. The remainder of the total variation for each data set may be due to parameters that are not addressed in this report, presumably due to a failure to identify them. However, given the limitations on the information that is provided by a survey of very low plutonium concentrations over irregular terrain, with sampling areas covering large areas within which considerable variation in concentrations is expected, it seems reasonable to assume that the unexplained variation, while relatively large, is due to random variation in the samples and their associated measurements. To some extent, the largest r^2 values are obtained from analysis of sectors where the plutonium concentrations were the highest and the total areas surveyed were the smallest; the smallest r^2 values are obtained from analysis of sectors where the plutonium concentrations were the lowest and the total areas surveyed were the largest.

CONCLUSIONS

An alternative approach to the analysis described in this report would be the use of multiple regression, taking the entire data set and analyzing it with respect to distance and direction from the center of the Rocky Flats Plant, in addition to the passage of time. Such a multiple regression would require a number of assumptions for which it may be difficult to obtain a consensus. For that reason, it was decided to isolate the data from each sector and perform a single regression of concentration against time, in the hope that at least part of the interpolation and future projections could be completed with a broad consensus about the validity of the assumptions that were used.

Analysis of the data with respect to distance and direction will be performed in the next phase of this project, using topographic projections to present the findings. While separation of the regression over time from analysis of the two-dimensional spatial distribution of concentrations presents some limitations to the strength of the final results, multiple regression need not be ruled out as a future project.

The findings of this report will not be finalized until after the 1991 surface soil survey is complete. At that time, the measured concentrations from the 1991 survey will be compared to the analysis described in this report, to see if they fall within the confidence bands of the predictive models. In addition, the 1991 survey results will be included in the regression analysis and the regression technique will be subject to modification, if indicated.

At the present time, it seems that it would be useful to composite all of the residuals to see if a larger residuals data set would turn up any flaws in the regression analysis that might have been overlooked to date.

Additional statistical tests of either the regression analysis or of the residuals may also be indicated.

The use of LLDs in the data sets may be a source of concern. A reevaluation of the applicability of LLDs in regression analysis, or an alternate treatment of LLDs, may be necessary.

REFERENCES

Co1977 Colorado Department of Health, *Radioactive soil contamination (Cesium-137 and Plutonium) in the environment near the Rocky Flats Nuclear Weapons Plant*, September 1977, revised April 1988.

P11981 Plane, Donald R., and Oppermann, Edward B., Business and Economic Statistics, *Chapter 13: Simple Regression* (Plano, Texas: Business Publications, Inc., 1981).

Dr1966 Draper, N.R., and Smith, H., Applied Regression Analysis, *Chapter 3: The Examination of Residuals* (New York: John Wiley & Sons, Inc., 1966).

Figure 1. Illustration of the steps used in performing an exponential fit of the data. Y values are transformed to lnY, the regression is performed, then calculated results are transformed back to their original units and plotted.

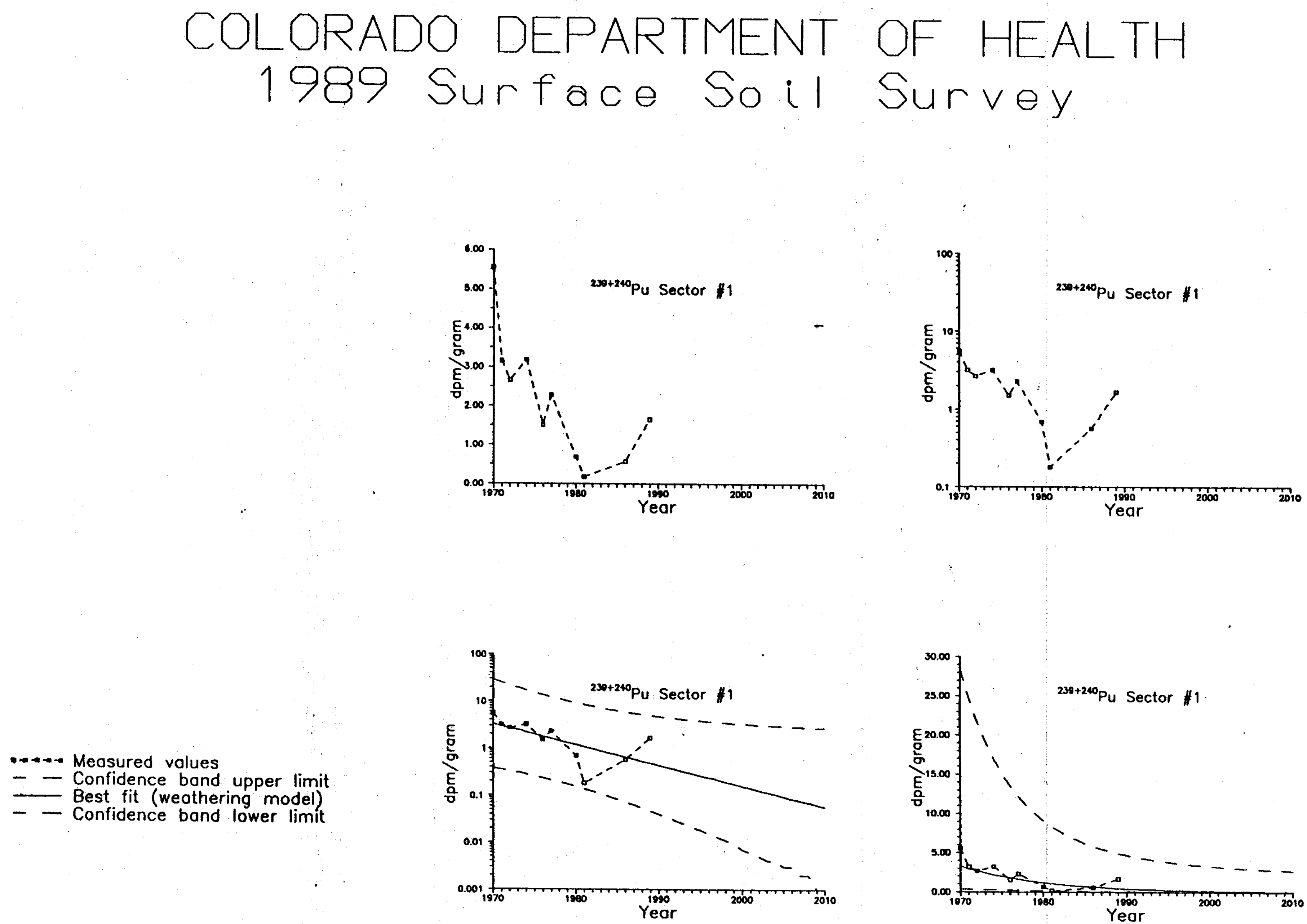


Figure 2. Illustration of the steps used in dividing a data set into subsets, then performing a linear regression on one data subset and an exponential fit on another. In this particular case, the measured plutonium concentration in 1970 was not employed in any regression, and the measured value remains as the best fitted data point for that year.

COLORADO DEPARTMENT OF HEALTH
1989 Surface Soil Survey

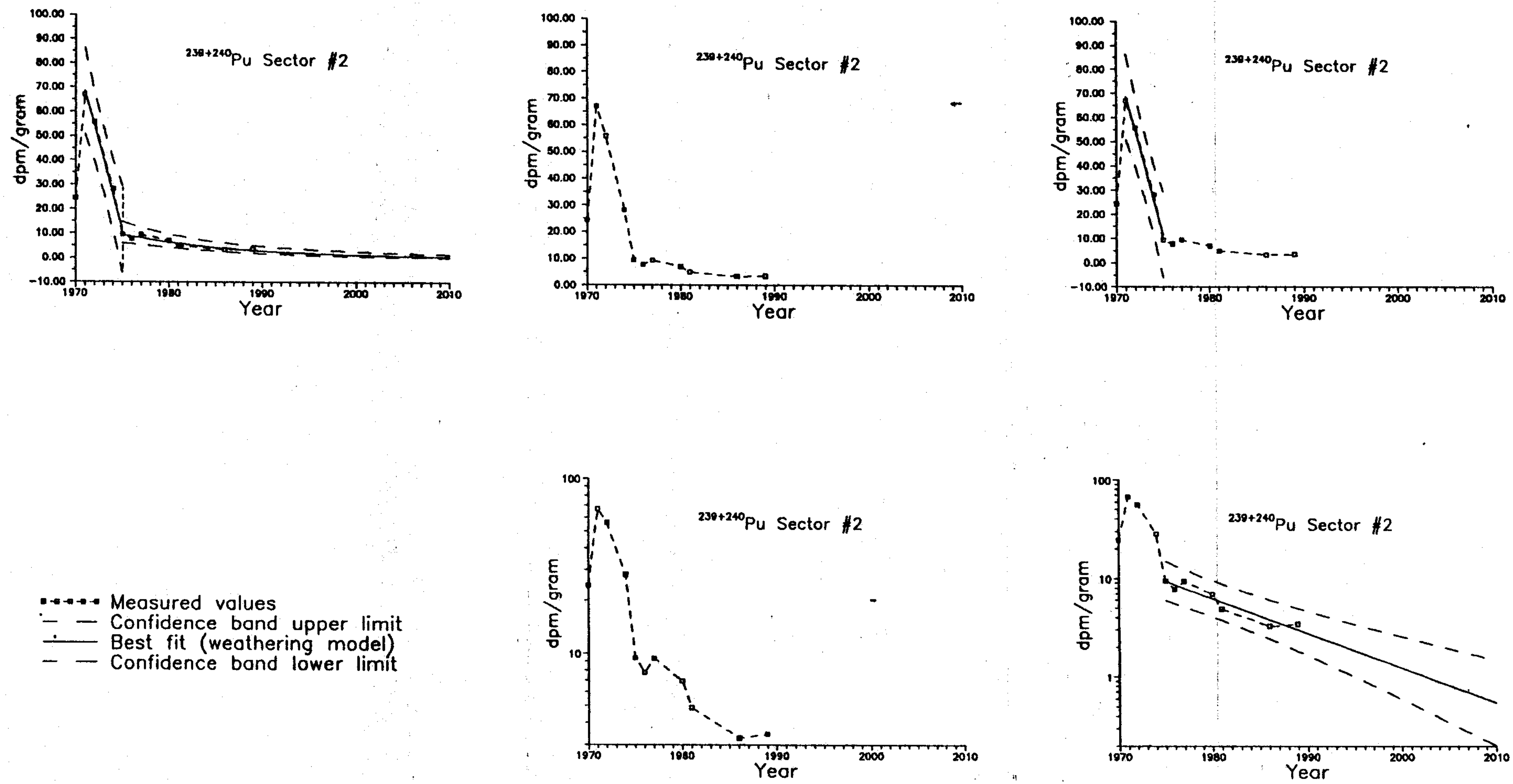


Figure 3. Summary of results of all regression analysis for the 13 sampling sectors near the Rocky Flats Plant.

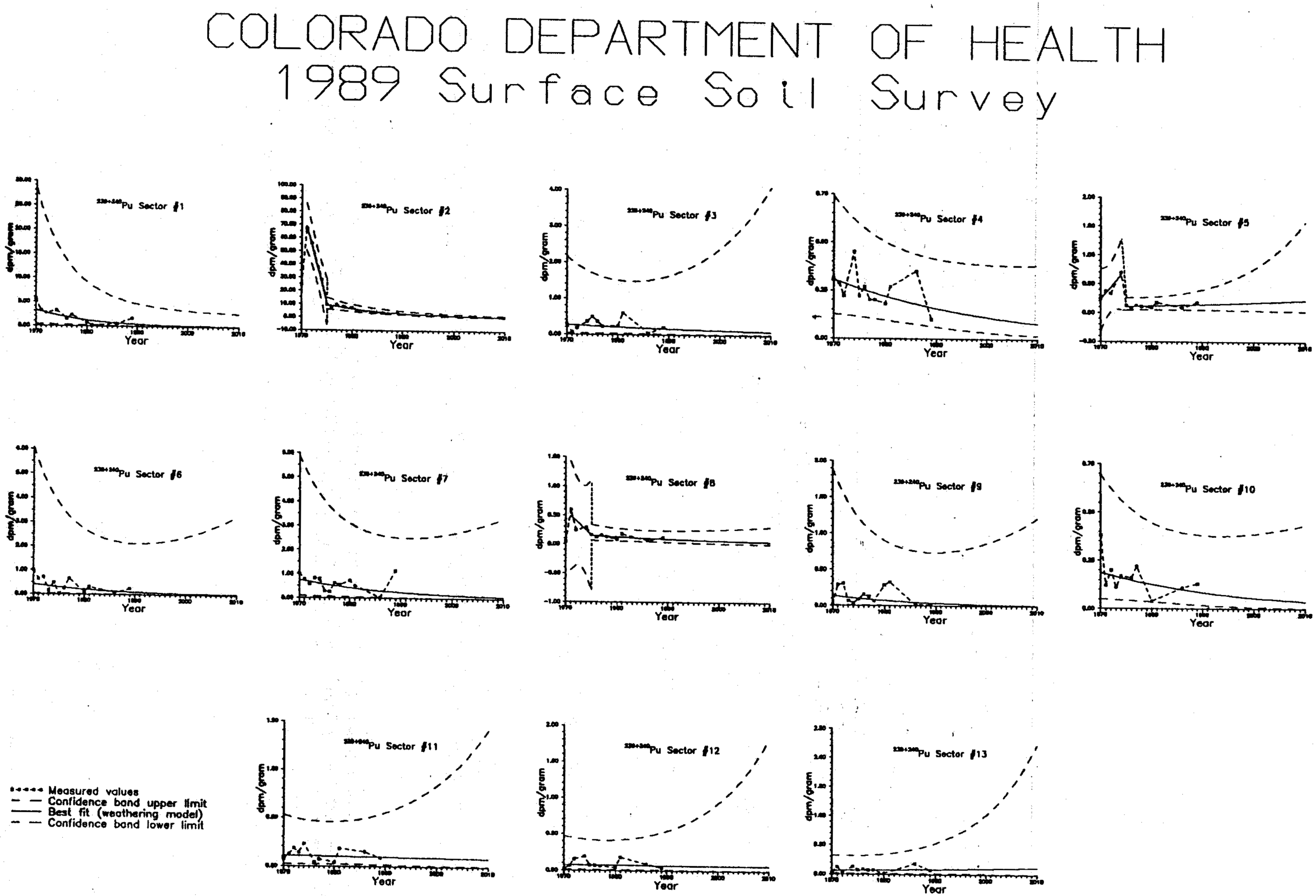


Figure 4. Summary of results of all regression analysis for the 8 sampling areas used to characterize plutonium contamination in surface soil that has resulted from worldwide fallout. Contamination in these areas is not presumed to have been transported directly from the Rocky Flats Plant.

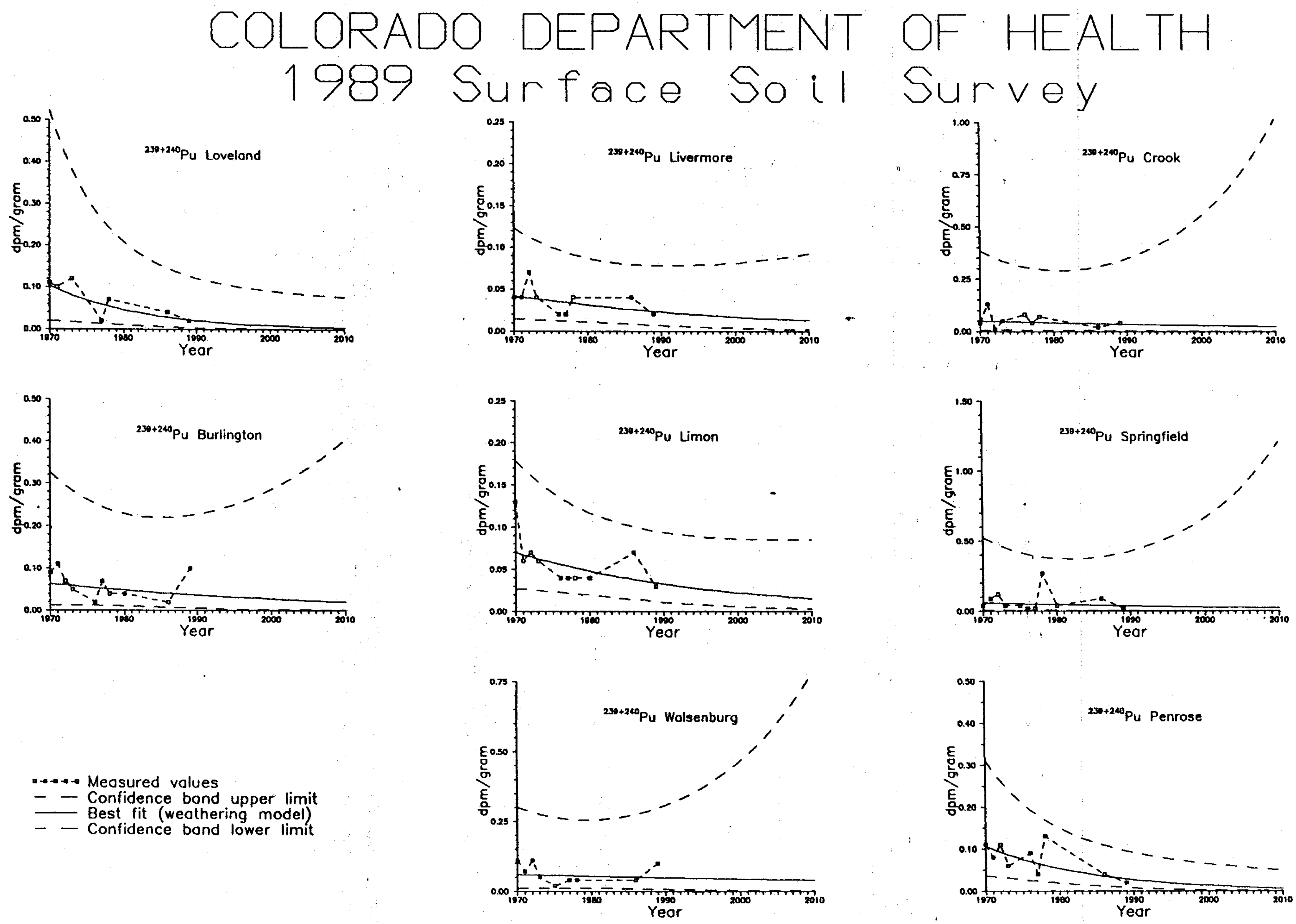
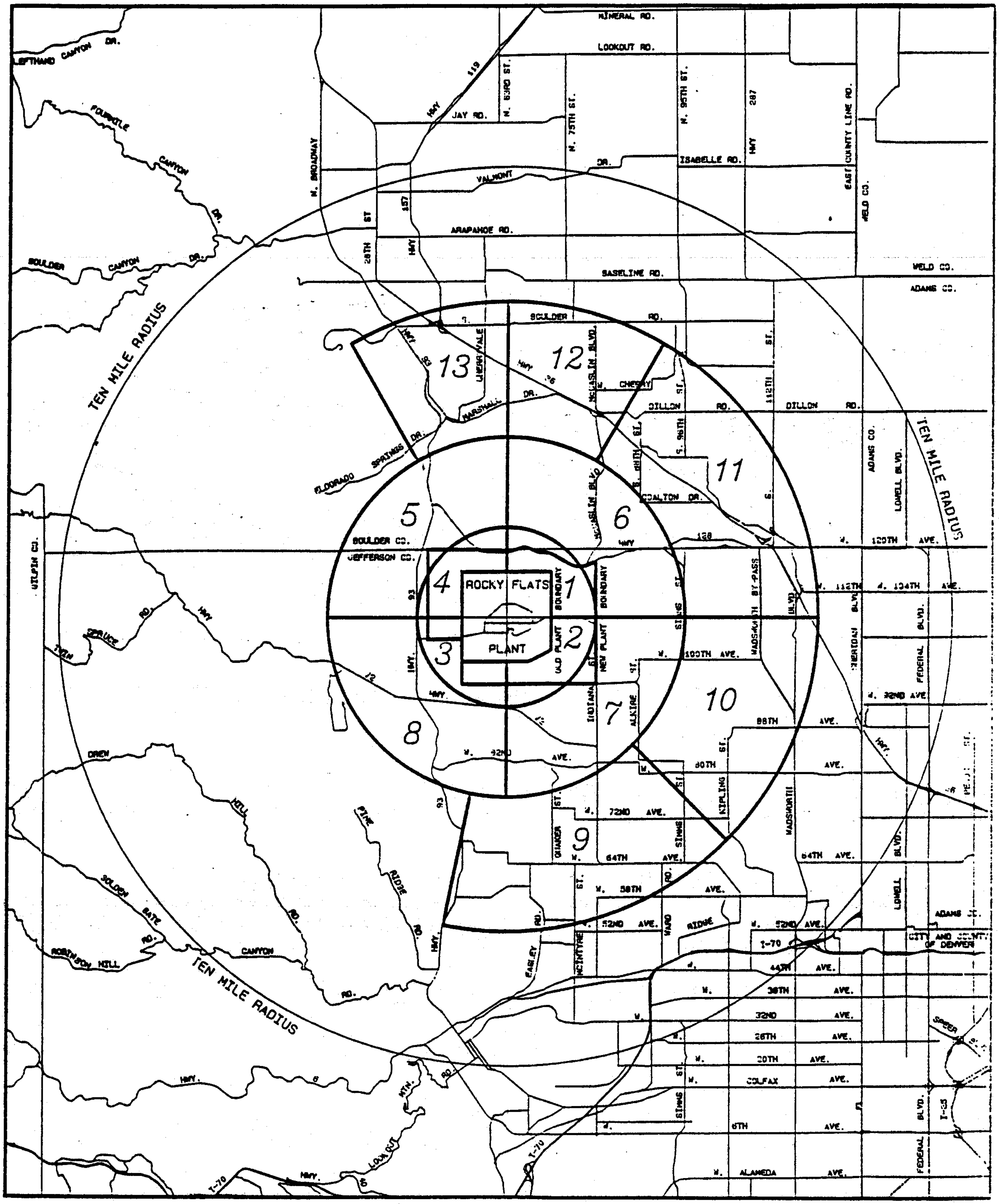


Figure 5.

ROCKY FLATS PLANT AREA - SOIL SECTOR IDENTIFICATION



MAPS PROVIDED BY: JEFFERSON COUNTY MAPPING DEPARTMENT

Figure 6. Residuals analysis from regression of data collected from the 13 sampling sectors near the Rocky Flats Plant.

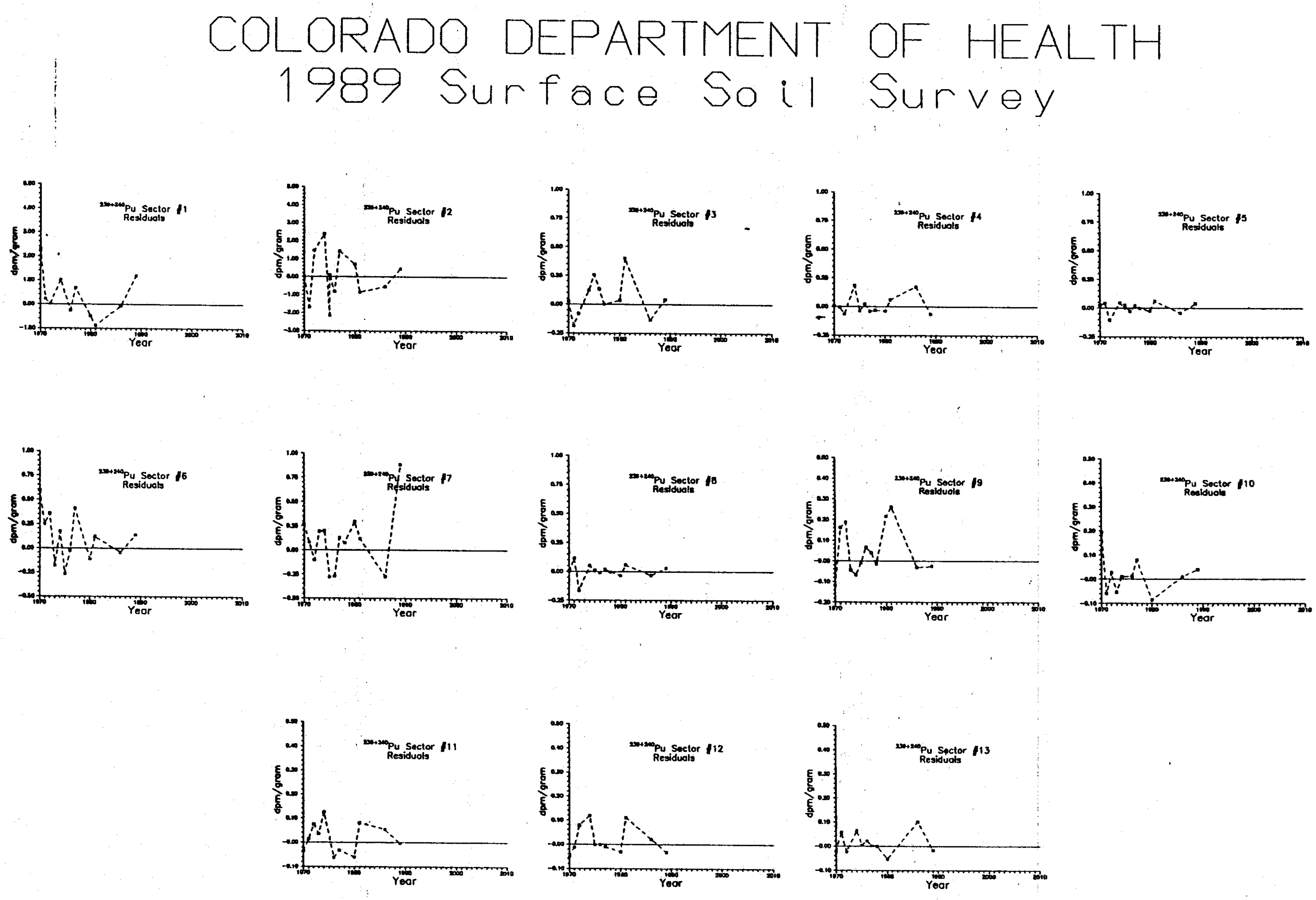


Figure 7. Residuals analysis from regression of data collected from the 8 sampling areas around Colorado that are not near Rocky Flats, but are subject to deposition of worldwide fallout.

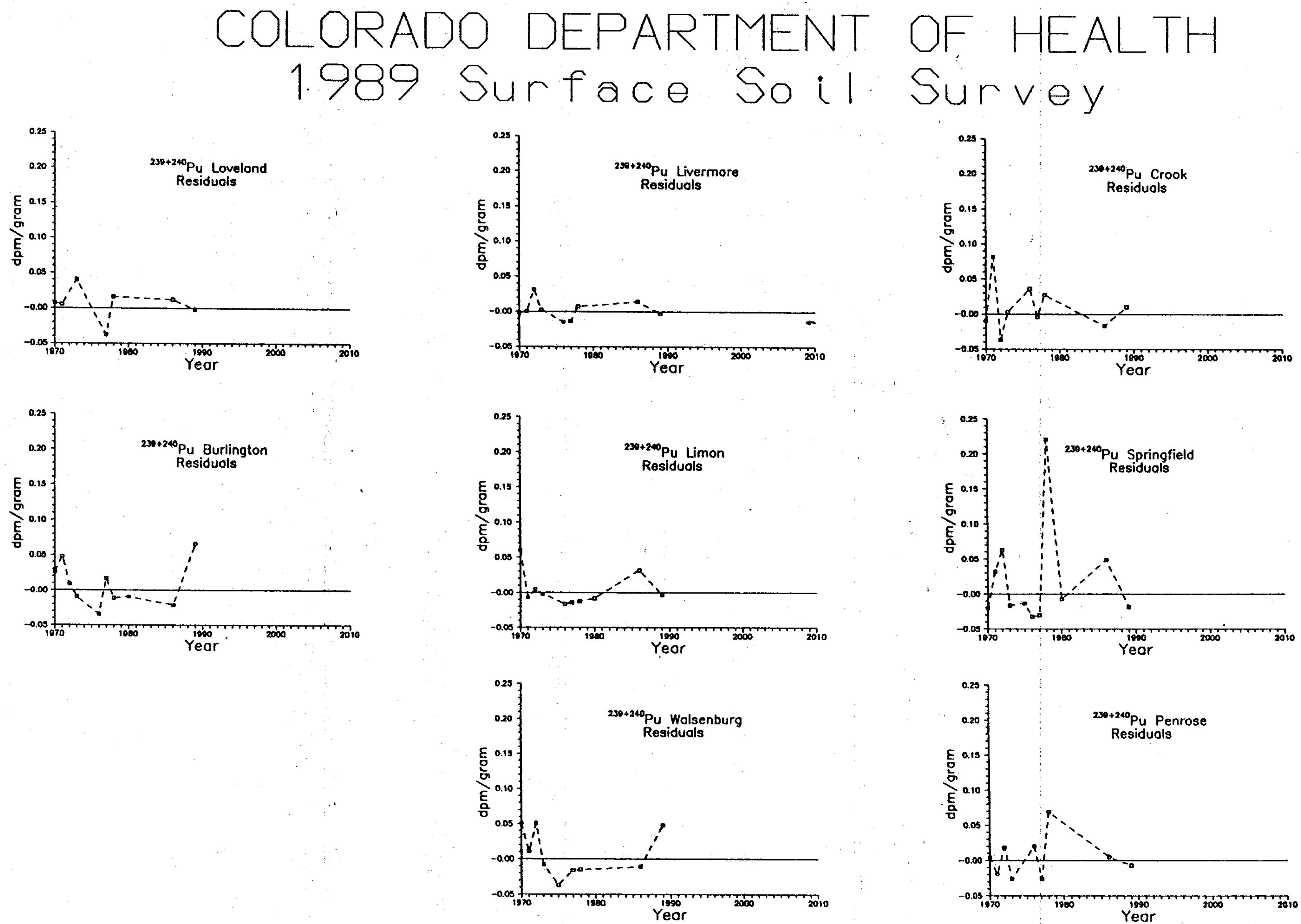


TABLE 1.

SUMMARY OF FORMULAS USED IN REGRESSION ANALYSIS OF PLUTONIUM CONCENTRATIONS OVER TIME, WITH CORRELATION COEFFICIENTS, STANDARD ERRORS OF THE ESTIMATES AND SUMS OF SQUARES (ERROR)

Sector 1	1970 –	$dpm/gm = e^{200.2601 - 0.10106 \times YEAR(A.D.)}$	$r^2 = .3991$ $S_{YX} = 0.8392$ $SSE = 5.6339$
Sector 2	1970	$dpm/gm = 24.40$	
	1971 – 1974	$dpm/gm = 28127.683 - 14.236 \times YEAR(A.D.)$	$r^2 = .9926$ $S_{YX} = 2.7558$ $SSE = 15.1887$
	1975 –	$dpm/gm = e^{162.1648 - 0.08098 \times YEAR(A.D.)}$	$r^2 = .9013$ $S_{YX} = 0.1543$ $SSE = 0.1190$
Sector 3	1970 –	$dpm/gm = e^{48.4921 - 0.02529 \times YEAR(A.D.)}$	$r^2 = .0366$ $S_{YX} = 0.8346$ $SSE = 6.2696$
Sector 4	1970 –	$dpm/gm = e^{65.0492 - 0.03362 \times YEAR(A.D.)}$	$r^2 = .2508$ $S_{YX} = 0.3548$ $SSE = 1.2591$
Sector 5	1970 – 1974	$dpm/gm = -215.9191 + 0.1097 \times YEAR(A.D.)$	$r^2 = .8705$ $S_{YX} = 0.0805$ $SSE = 0.0157$
	1975 –	$dpm/gm = e^{-43.7787 + 0.02103 \times YEAR(A.D.)}$	$r^2 = .1103$ $S_{YX} = 0.3193$ $SSE = 0.6116$
Sector 6	1970 –	$dpm/gm = e^{134.7906 - 0.06888 \times YEAR(A.D.)}$	$r^2 = .1321$ $S_{YX} = 1.1027$ $SSE = 12.1590$
Sector 7	1970 –	$dpm/gm = e^{107.5249 - 0.05473 \times YEAR(A.D.)}$	$r^2 = .1276$ $S_{YX} = 0.8532$ $SSE = 8.0076$
Sector 8	1970	$dpm/gm = 0.04$	
	1971 – 1974	$dpm/gm = 164.0815 - 0.0830 \times YEAR(A.D.)$	$r^2 = .6191$ $S_{YX} = 0.1456$ $SSE = 0.0424$
	1975 –	$dpm/gm = e^{60.0366 - 0.03137 \times YEAR(A.D.)}$	$r^2 = .2404$ $S_{YX} = 0.2981$ $SSE = 0.5333$
Sector 9	1970 –	$dpm/gm = e^{120.7739 - 0.06231 \times YEAR(A.D.)}$	
Sector 10	1970 –	$dpm/gm = e^{31.6757 - 0.01691 \times YEAR(A.D.)}$	$r^2 = .1836$ $S_{YX} = 0.5623$ $SSE = 2.5291$
Sector 11	1970 –	$dpm/gm = e^{10.0050 - 0.006211 \times YEAR(A.D.)}$	$r^2 = .0041$ $S_{YX} = 0.6338$ $SSE = 3.6154$
Sector 12	1970 –	$dpm/gm = e^{2.3933 - 0.002548 \times YEAR(A.D.)}$	$r^2 = .0049$ $S_{YX} = 0.7393$ $SSE = 4.9187$
Sector 13	1970 –	$dpm/gm = e^{-33.0925 + 0.01539 \times YEAR(A.D.)}$	$r^2 = .0222$ $S_{YX} = 0.6448$ $SSE = 3.7417$
Loveland	1970 –	$dpm/gm = e^{157.9332 - 0.08132 \times YEAR(A.D.)}$	$r^2 = .5856$ $S_{XY} = 0.5502$ $SSE = 1.5135$
Livermore	1970 –	$dpm/gm = e^{55.9464 - 0.03002 \times YEAR(A.D.)}$	$r^2 = .2122$ $S_{XY} = 0.4106$ $SSE = 1.1800$
Crook	1970 –	$dpm/gm = e^{33.5214 - 0.01854 \times YEAR(A.D.)}$	$r^2 = .0281$ $S_{XY} = 0.7748$ $SSE = 4.2026$
Burlington	1970 –	$dpm/gm = e^{51.8660 - 0.02772 \times YEAR(A.D.)}$	$r^2 = .0797$ $S_{XY} = 0.6334$ $SSE = 3.2100$
Limon	1970 –	$dpm/gm = e^{71.7906 - 0.03779 \times YEAR(A.D.)}$	$r^2 = .3244$ $S_{XY} = 0.3667$ $SSE = 1.0756$
Springfield	1970 –	$dpm/gm = e^{44.4143 - 0.02398 \times YEAR(A.D.)}$	$r^2 = .0300$ $S_{XY} = 0.8697$ $SSE = 6.8081$
Walsenburg	1970 –	$dpm/gm = e^{17.2493 - 0.01019 \times YEAR(A.D.)}$	$r^2 = .0137$ $S_{XY} = 0.6149$ $SSE = 2.6466$
Penrose	1970 –	$dpm/gm = e^{132.3771 - 0.06834 \times YEAR(A.D.)}$	$r^2 = .5860$ $S_{XY} = 0.4079$ $SSE = 1.1644$